REMARKS

Claim 1 has been amended to overcome the rejection thereof under 35 USC 112, paragraph 2. Claim 8 has been similarly amended. Claim 9 has been amended, as suggested on page 3 of the office action, to overcome the rejection thereof under 35 USC 101. New claims 10 and 11 have been added. Claims 10 and 11 respectively depend on claims 8 and 9, and include the same limitations as set forth in claim 5, which the office action indicates contains allowable subject matter.

Applicant traverses the rejection of claim 9 under 35 USC 112, second paragraph as being indefinite. The office action states it is unclear as to whether a storage medium or a storage device is being claimed. In fact, claim 9 is directed to both a storage medium and a storage device. A storage medium is, for example, a magnetic tape, while a storage device is, for example, a semiconductor memory. MPEP Section 2173.05(h) indicates alternative expressions using "or" are acceptable. This portion of the MPEP specifically indicates, for example, that phraseology such as "iron, steel or any other magnetic material" is acceptable. If the rejection is repeated, the examiner is requested to indicate more specifically why claim 9 does not comply with 35 USC 112, second paragraph.

Applicant traverses the rejection of claim 9 under 35 USC 112, first paragraph, as failing to comply with the written description requirement. The office action alleges the limitation in claim 9 for "storage device or storage medium storing coded indicia" is not found in the specification as filed. The storage device is based on the statement in the specification at paragraph [0029] of applicant's published application at which states "a specific application implemented in a personal computer;" and "a language identification device DI according to the invention whose main software components relating to automatic language identification." Withdrawal of the rejection is in order.

Applicant traverses the rejection of claims 1-4 and 6-9 as being unpatentable over Van den Akker, US Patent 6,415,250, in view of De Campos, US Patent 6,272,456.

Van den Akker discloses an automatic language identification system 110 (Fig. 3) based on a probability analysis of predetermined word portions extracted from input text 301, the language of which is to be identified. A word portion is defined as the portion of a

word at the end of a word having a predetermined number of characters (column 11, line 65, through column 12, line 7, and column 20, lines 4-9), which is generally a suffix, and the portion of a word at in the beginning of a word in the form of a prefix (column 8, line 45, through column 9, line 3; Fig. 2C).

Language corpus analyzer 302 associates each word portion of a predetermined language corpus 309 with a normalized frequency indicative of the number of times the word portion is found within the corpus (column 9, lines 35-42, and column 12, lines 44-60) and with a relative probability derived from the frequency in relation to the size of the corpus. When the word portion rarely appears in the language corpus, the probability is close to zero (column 10, lines 56-60, and column 13, liens 38-41).

Language identification engine 306 sums the relative probability values associated with each language for each of the word portions extracted from input text 301 and found in probability table 304. Engine 306 identifies the language of the input text 301 having the largest total accumulated relative likelihood value (column 10, lines 33-45). As a result, the identification language system 110 relates to one category of first character strings (suffixes or prefixes) in a word (column 20, liens 52, 53, 66 and 67; claim 1).

Analyser 302 of Van den Akker analyzes only one character string per extracted word with respect to language corpus 309, whereas the analyzing means or step in applicant's claims 1 and 8 analyzes plural character strings for an extracted word. Analyser 302 of Van den Akker thus fails to carry out the function of applicant's claimed analyzing means or step.

Moreover, system 110 of Van den Akker applies to each found character string a probability dependent on the frequency of the found character string in a language corpus and not the location of the found character string in the word extracted from the input text. However, as paragraph [0016] of Applicant's published application indicates, the present invention is not limited to trigrams or word portions having a particular location. In the analyzing means and step of claims 1 and 8, "all character strings contained in said extracted word" stands for all character strings "having lengths lying between one character and the number of characters in said extracted word" and including the chains that are partially overlapping (see [0057] of applicant's published application).

The atypical character strings in Van den Akker are associated with a probability value of zero or a common negative value: -0,99. Applicant's claims 1 and 8 require a score associated with the predetermined language , i.e., with said each determined language, to increase by a first coefficient depending on the position of the first character string found in said extracted word. Whenever a second character string is found in the extracted word, the score decreases by a respective second coefficient that is associated with said found second character string. The respective second coefficient is required to increase as the probability of said found second character string in said each determined language decreases. Hence, the second coefficient is different for each of the atypical character strings, resulting in significantly improved language identification accuracy. The language identification system of Van den Akker is far less accurate than that of the system defined by claim 1 or the method defined by claim 9.

Therefore, Van den Akker fails to disclose the analyzing means and the various comparing means of applicant's independent claim 1 or the corresponding steps of applicant's claim 8. Furthermore, Van den Akker does not disclose means for or the step of prestoring "first" frequently character strings and a means for or step of prestoring "second" atypical character strings before analyzing extracted words and comparing the strings contained in extracted words.

The Examiner admits Van den Akker does not disclose that whenever a second character string is found in said extracted word, said score is decreased by a respective second coefficient (that is associated with said found second character string), said respective second coefficient increasing as the probability of said found second character string in said each determined language decreasing. Applicant does not agree with the conclusion in the office action that one of ordinary skill in the art would have modified Van den Akker as a result of De Campos to increase and decrease the second coefficient as required by claims 1 and 8.

De Campos identifies a language of a document by using n-gram profiles, such as 3-gram profiles, 4-gram profiles and 5-gram profiles (column 3, lines 1-8). A longest match is kept if a letter window within the sample input, such as a sentence, contains one or more matches as a result of a comparison with the reference letter sequences in the n-

gram profiles for each language (column 3, lines 8-14).

Referring to the passage in column 3, lines 60-67, cited in page 8 of the Office Action, De Campos discloses a first condition for increasing the score for one of the languages: the frequency parameters of the longest match indicate the longest match is found in relatively few of the languages. Increasing the score for a language thus depends on the presence of the longest match in other languages. This means the score increases when the probability to find the longest match increases in plural languages.

In claims 1 and 8, the score is decreased, not for the "longest match," but whenever a second character string is found in the extracted word for one predetermined language. Thus the score decrease of claims 1 and 8 is not related to (1) the length of the second character string or (2) the presence of a second character string in other languages.

De Campos discloses a second condition for decreasing the score for one of the languages: the frequency parameters of the longest match indicate the longest match is found in many languages. This means the score decreases when the probability to find the longest match increases in many languages. Claims 1 and 8 state another condition: the score associated with a determined language decreases by a second coefficient associated with a found second character string and the second coefficient increases as the probability of the found second character string in said determined language decreases.

De Campos fails to disclose second coefficients that are respectively associated with second fixed character strings in dependence on a predetermined language. The relied on portion of De Campos, i.e., column 3, lines 60-67, fails disclose a second coefficient for both a language and a character.

Therefore, the language identification system of claim 1 and the method of claim 8 are not made obvious by the combination of Van den Akker and De Campos.

Dependent claims 2-4, 6 and 7 are allowable for the same reasons advanced for claim 1, upon which they depend.

Allowance is in order.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of

this paper, including extension of time fees, to Deposit Account 07-1337 and please credit any excess fees to such deposit account.

Respectfully submitted,

LOWE HAUPTMAN HAM & BERNER, LLP

/Allan M. Lowe/

Allan M. Lowe
Registration No. 19,641

1700 Diagonal Road, Suite 300
Alexandria, Virginia 22314
(703) 684-1111
(703) 518-5499 Facsimile
Date: October 22, 2008
AML/cjf